ADVANCED SPEECH ENCODING PROGRAM: RELATED REFERENCES

NOTE: The following is a list of references that may be pertinent to the DARPA Advanced Speech Encoding Program. It is not meant to be all inclusive.

<u>GENERAL SPEECH BOOKS</u>

L.R. Rabiner and R.W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1978.

T.F. Quatieri, Discrete-Time Speech Signal Processing: Principles and Practice, Prentice-Hall, Inc., Englewood Cliffs, NJ, 2002.

D. O'Shaughnessy, Speech Communication: Human and Machine, Addison-Wesley, Reading, MA, 1987.

B. Gold and N. Morgan, Speech and Audio Signal Processing, John Wiley and Sons, Inc., 2000.

J.R. Deller, J.G. Proakis, and J.H.L. Hansen, Discrete-time Processing of Speech, Macmillan Publishing Co., New York, NY, 1993.

S. R. Quackenbush, T.P. Barnwell, and M.A. Clements, Objective Digital Signal Processing, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1988.

L. Rabiner and B.H. Juang, Fundamentals of Speech Recognition, Prentice-Hall, Englewood Cliffs, NJ, 1993.

Special Issue: NIST 1999 Speaker Recognition Workshop, Eds. J.P. Campbell and J. Schroeder, Academic Press, vol. 10, no. 1-3, January/April/July 2000.

Speech Coding and Synthesis, Edited by W.B. Kleijn and K.K. Paliwal, Elsevier, 1995.

J.L. Flanagan, Speech Analysis, Synthesis, and Perception, Springer-Verlag, Second Ed., New York, NY, 1972.

<u>EM RADAR SENSOR</u> (from http://speech.llnl.gov/)

Ng L.C., Holzichter, J. F., Larson, P.,, " *Low Bandwidth Vocoding Using EM Sensor and Acoustic Signal Processing*", Abstract 2001 for ICASSP Salt Lake City, 2001

Holzrichter, Ng, Burke, Kobler, and Rosowski , *"<u>A description of EM wave interactions with the glottis and vocal tract walls</u>"*, Poster presented at the Acoustics Society Meeting, Cancun, Mexico Dec 1-6, 2002

Holzrichter, Ng, Burke, Champaign at LLNL, Kobler, Rosowski, and Hillman at Mass. Eye and Ear Infirmary., Lawrence Livermore Laboratory and UC Davis report: UCRL-JRNL-147775 - PDF Format . Accepted for publication JASA 2004. "Measurements of Glottal Structure Dynamics".

Hernstein at LLNL and UCD. This study examines the statistics of the time duration of unvoiced speech units that usually precede or follow voiced speech units UCRL-TR-155600 - PDF Format

Titze, I.R., Story, B.H., Burnett, G.C., Holzrichter, J.F., Ng, L.C., and Lea, W.A., "Comparison between electroglottography and electromagnetic glottography", JASA, 107(1), January 2000.

Burnett, G. C.; Holzrichter, J. F.; Gable, T. J.; and Ng, L. C., "Denoising of Human Speech using Combined Acoustic and EM Sensor Signal Processing," presentedat the International Conference on Acoustics, Speech, and Signal Processing--ICASSP 2000, June 5-9, 2000, Istanbul, Turkey. Conference Paper - pdf format.

Burnett, G. C.; Holzrichter, J. F.; Gable, T. J.; and Ng, L. C., "The Use of Glottal Electromagnetic Micropower Sensors (GEMS) in Determining a Voiced Excitation Function, " presented at the 138th Meeting of the Acoustical Society of America, November 2, 1999, Columbus, Ohio.

Gable, T. J.; Burnett, G. C.; Holzrichter, J. F.; and Ng, L. C., "Speaker Verification Performance Comparison Based on Traditional and Electromagnetic Sensor Pitch Extraction, " presented at the 138th Meeting of the Acoustical Society of America, Poster Session 2pSC, Speech Processing, November 2, 1999, Columbus, Ohio.

Ng, L. C.; Burnett, G. C.; Holzrichter, J. F.; and Gable, T. J., "Background Speaker Noise Removal Using Combined EM Sensor/Acoustic Signal Signals, "presented at the 138th Meeting of the Acoustical Society of America, Poster Session 2pSC, Speech Processing, November 2, 1999, Columbus, Ohio.

Burnett, G. C.; Holzrichter, J. H., "Human Speech Articulator Measurements Using Low Power, 2 GHz Homodyne Sensors", presented at the International Conference on Infrared and Millimeter Waves, September 6, 1999, Monterey, CA.

Dr. Burnett's thesis, entitled "The Physiological Basis of Glottal Electromagnetic Micropower Sensors (GEMS) and Their Use in Defining an Excitation Function for the Human Vocal Tract", was the recipient of the 1998-99 U.C. Davis Allen G. Marr Prize for the outstanding thesis in the colleges of Mathematics, Physical Science, and Engineering.

Burnett, G. C.; Holzrichter, J. F.; Gable, T. J.; and Ng, L. C., "Direct and Indirect Measures of Speech Articulator Motions Using Low Power EM Sensors," presented at the XIV International Congress of Phonetic Sciences, August 6, 1999, San Francisco, California.

Burnett, G.C., Holzrichter, J.F., Cheng, C.Z., Gable, T.J., Ng, L.C., Freeman, R.R., (1999)"Speech Characterization Using Micro-Power EM Sensors" presented at the NSF Workshop on the Human-Computer Interaction Program, February 22, 1999, Orlando, Florida.

Burnett, G.C., Gable, T. J., Holzrichter, J.F., Ng, L.C., (1998)"Micropower Electro-Magnetic Sensors for Speech Characterization" presented at the International Conference and Speech and Linguistic Phonetics, November 30 - December 4, 1998, Sydney Australia.

Holzrichter, J. F., Lea, W. A., McEwan, T. E., Ng, L. C., Burnett, G. C. (1996). "Speech Coding, Recognition, and Synthesis using Radar and Acoustic Sensors," University of California Report UCRL-ID-123687 (reprints available from the Office of Scientific and Technical Information, P.O.Box 62, Oak Ridge, TN, 37831).

Burnett, G.C., Holzrichter, J.F., Ng, L.C., Leonard, R., Lea, W.A. (1996) "Micropower Radar Measurements of Human Vocal Articulator Motions" Third Joint Meeting of the Acoustical Society of America and Acoustical Society of Japan, December 2-6, 1996, Honolulu, Hawaii.

Holzrichter, J.F., Lea, W.A., Ng, L.C., Burnett, G.C., (1996)"Uses of Micropower Radars for Speech Coding and Applications" Third Joint Meeting of the Acoustical Society of America and Acoustical Society of Japan, December 2-6, 1996, Honolulu, Hawaii.

Holzrichter, J.F. and Burnett, G.C. (1997) "Use of Low Power EM Radar Sensors for Speech Articulator Measurements" presented at Eurospeech 97, September 22-25, 1997, Rhodes, Greece and published in proceedings.

Gable, T. J., Burnett, G.C., Holzrichter, J.F., Ng, L.C., Lea, W.A. (1997) "Comparison of conventional acoustic and MIR radar/acoustic processing of speech signals" paper 4aSP5, 134th meeting of the Acoustic Society of America, December 1-5, 1997 San Diego.

Burnett, G.C., Gable, T. J., Holzrichter, J.F., Ng, L.C., (1997)"Voiced excitation functions calculated from micropower impulse radar information." paper 4aSP4. 134th meeting of the Acoustic Society of America, December 1-5, 1997 San Diego.

Holzrichter, J.F., Burnett, G.C., Ng, L.C., Lea, W.A. (1998)"Speech articulator measurements using low power EM-wave sensors." Journal of the Acoustic Society of America, 103 (1), pg. 622-625, January 1998. Reprints available upon request.

Haddad, W. S., Rosenbury, E. T., Johnson, K. B., and Pearce, F. J. (1997) "Measurements of the Dielectric Properties of Body Tissues and Fluids at Microwave frequencies," to be published, Lawrence Livermore National Laboratory.

Holzrichter, J.F., Ng, L.C. (1998), Patent Number 5,729,694, "Speech Coding, Reconstruction and Recognition Using Acoustics and Electromagnetic Waves."

Gable's doctoral thesis on "Speaker verification using GEMS", UC Davis.

SPEAKER RECOGNITION AND EVALUATIONS

J. P. Campbell, "Speaker recognition: A Tutorial," Proceedings of the IEEE, vol. 85, pp. 1437--1462, September 1997.

H. Gish and M. Schmidt, "Text-independent speaker identification", IEEE Signal Processing Magazine, vol. 11, no. 4, pp. 18--32, October 1994.

G. Doddington, M. Przybocki, A. Martin, and D. A. Reynolds, "The NIST speaker recognition evaluation - overview, methodology, systems, results, perspective," Speech Communication, June 2000.

D.A. Reynolds, "Speaker identification and verification using Gaussian mixture speaker models", Speech Communication, vol. 17, pp. 91-108, August 1995.

A. Schmidt-Nielsen and T. Crystal, ``Speaker verification by human listeners: Experiments comparing human and machine performance using the NIST 1998 speaker evaluation data", Digital Signal Processing, Special Issue: NIST 1999 Speaker Recognition Workshop, Academic Press, vol. 10, no. 1-3, pp. 249--266, January/April/July 2000.

NIST, ``NIST speaker recognition workshop notebook", NIST administered speaker recognition evaluation on the Switchboard corpus, 1996-2002.

F. Soong, A. Rosenberg, L. Rabiner, and B. Juang, ``A vector quantization approach to speaker recognition", Proc. Int. Conf. Acoustics, Speech, and Signal Processing, vol. 1, pp. 387--390, Tampa, FL, 1985.

F. Soong and A. Rosenberg, ``On the use of instantaneous and transitional spectral information in speaker recognition, IEEE Trans. Acoustics, Speech, and Signal Procesing, vol. ASSP-36, no. 6, pp. 871-879, June 1988.

T.F. Quatieri, C.R. Jankowski, and D.A. Reynolds, "Energy onset times for speaker identification", IEEE Signal Processing Letters, vol. 1, no. 11, pp. 160--162, November 1994.

[LDC] Linguistic Data Consortium, http://www.ldc.upenn.edu/ldc/catalog/html/speech-html.

GLOTTAL FLOW ESTIMATION FROM THE WAVEFORM

M. Plumpe, T.F. Quatieri, and D.A. Reynolds, "Modeling of the glottal flow derivative waveform with application to speaker identification", IEEE Trans. Speech and Audio Processing, vol. 1, no. 5, pp. 569-586, September 1999.

D.Y. Wong, J.D. Markel, and A.H. Gray Jr, "Least squares glottal inverse filtering from the acoustic speech waveform", IEEE Trans. Acoustics, Speech, and Signal Processing, vol. ASSP-27, no. 4, pp. 350--355, August 1979.

B.S Atal, ``Automatic recognition of speakers from their voices", IEEE Proceedings, vol. 64, no. 4, pp. 460--475, 1976. (First use of pitch in speaker recognition.)

J. He, L. Liu, and G. Palm, "On the use of features from prediction residual signals in speaker identification", Proc. Eurospeech95, vol. 1, pp. 313--316, Madrid,
Spain, 1995.

P. The venaz and H. Hugli, "Usefulness of the {LPC}-residue in text-independent speaker verification", Speech Communication, vol. 17,
no. 1-2, pp. 145--157, August 1995.

L.R. Yanguas, T.F. Quatieri, and F. Goodman, "Implications of glottal source for speaker and dialect identification", Proc. Int. Conf. Acoustics, Speech, and Signal Processing, vol. 2, pp. 813--816, Phoenix, Arizona, 1999.

T.F. Quatieri, C.R. Jankowski, and D.A. Reynolds, "Energy onset times for speaker identification", IEEE Signal Processing Letters, vol. 1, no. 11, pp. 160--162, November 1994.

K.E. Cummings and M.A. Clements, "Analysis of glottal waveforms across stress styles", Proc. Int. Conf. Acoustics, Speech, and Signal Processing, pp. 369--372, Albuquerque, NM, 1990.

D.G. Childers and C. Ahn, "Modeling the glottal volume-velocity waveform for three voice types," J. Acoustical Society of America, vol. 97, no. 1, pp. 505--519, January 1995.

T.V. Ananthapadmanabha and G.Fant, "Calculation of true glottal flow and its components", Speech Communications, vol. 1, pp. 167--184, 1982.

G.Fant, "Glottal flow: models and interaction", J. Phonetics, vol. 14, pp. 393--399, 1986.

MULTIPLE SENSORS FOR SPEECH

Viswanathan, et. al, ICASSP 84, "Multisensor Speech Input for Enhanced Immunity to Acoustic Background Noise." Throat accelerometer plus mike shows enhanced intelligibility and quality in cockpit noise.

Viswanathan, et. al., ICASSP 85, "Noise-Immune Speech Transduction using Multiple Sensors." Continuation of work reported in 84, with more conditions and sensor placements.

Roucos, Viswanathan, et. Al., ICASSP 86, "Word recognition using Multisensor Speech Input in High Ambient Noise." Shows improved word recognition in noise with combined sensors. Note that both training and test data collected in controlled multisensor configuration.

Harrison, Lim, and Singer, ICASSP 84, "Adaptive Noise Cancellation in a Fighter Cockpit Environment," Shows noise cancellation with one mike outside facemask and one mike inside facemask.

R. P. Lippmann, "Detecting Nasalization using a Low-Cost Miniature Accelerometer," j. of speech & Hearing research, vol. 24, pp 314-317, Sept. 1981. Aimed at evaluating nasalization in deaf talkers.

J.M. Kates, "An evaluation of hearing aid array processing", IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics", 1995, New Paltz, New York, p. 15-18, October 1995.

S. Nandkumar and J.H. Hansen, "Dual-channel iterative speech enhancement with constraints on an auditory-based spectrum", IEEE Trans. Speech and Audio Processing, vol. 3, no. 1, Jan. 1995. (See also B. Widrow and S.D. Stearns, Adaptive Signal Processing,Prentice Hall, Englewood Cliffs, NJ, 1985, for a variety of two-channel speech enhancement scenarios.)

AUGMENTING SPEECH WITH VIDEO LIPREADING

G. Gravier, G. Potamianos, and C. Neti, "Asynchrony modeling for audio-visual speech recognition", Proc. Human Language Technology Workshop, sponsored by DARPA, NSF, and ACL, 2002.

E.D. Petajan, "Automatic Lipreading to Enhance Speech recognition," IEEE Globecomm, 1984 (also Petajan's PhD thesis)

Tomlinson, Russell, and Brooke, "Integrating Audio and Visual Information to Provide highly Robust Speech Recognition," ICASSP-96, pp. 821-824

Meier, Hurst, and Duchnowski, "Adaptive Bimodal Sensor Fusion for Automatic Speechreading," ICASSP-96, pp. 833-836

V. Lesbros, "From images to sounds: A dual representation", Computer Music Journal, Fall 1996.

SPEECH CODING

A. Kondoz, Digital Speech: Coding For Low Bit Rate Communication Systems, John Wiley & Sons, New York, 1994.

N.S. Jayant and P. Noll, Digital Coding of Waveforms: Principles and Applications to Speech and Video, Prentice-Hall, Englewood Cliffs, New Jersey, 1984.

A. Gersho and R. Gray, Vector Quantization and Signal Compression, Dordrecht, Holland: Kluwer Academic Publishers, 1991.

J.P. Campbell, Jr., T.E. Tremain, and V.C. Welch, "The federal standard 1016 4800 bps CELP voice coder", Digital Signal Processing, Academic Press, vol. 1, no. 3, pp. 145--155, 1991.

A. Gersho, "Advances in speech and audio compression, "Proceedings of The IEEE, vol. 82, pp. 900-918, June 1994.

Spanias, A. S. (1994). Speech Coding: A Tutorial Review. Proceedings of the IEEE, Oct. 1994

[J. Schroeter and M.M. Sondhi, ``Speech coding based on physiological models of speech production'', Chapter in Advances in Speech Signal Processing, Edited by S. Furui and M.M. Sondhi, Marcel Dekker, New York, NY, 1992.

T.F. Quatieri, R.B. Dunn, D.A. Reynolds, J.P. Campbell, and E. Singer, ``Speaker Recognition using G.729 Speech Codec Parameters'', IEEE Int. Conf. Acoustics, Speech, and Signal Processing, vol. 2., pp. 1089-1093, Istanbul, Turkey, June 2000.

R. Salami et al., ``Design and description of CS-ACELP: A toll quality 8 kb/s speech coder'', IEEE Trans. on Speech and Audio Processing, vol. 6, no. 2, pp. 116--130, March 1998.

M.R. Schroeder and B.S. Atal, "Code-excited linear prediction (CELP): High-quality speech at very low bit rates", Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, vol. 3, pp. 937--940, April 1985.

A. McCree, K. Truong, E.B. George, T.P. Barnwell, and V, Viswanathan, "A 2.4 kbit/s MELP coder candidate for the new US Federal standard", Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, pp.~202--203, 1996.

A. McCree, "A 4.8 kbit/s MELP coder candidate with phase alignment", Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, Istanbul, Turkey, vol. 3, pp. 1379--1382, June 2000.

ITU-T Recommendation G729, "Coding of speech at 8 kb/s using conjugate-structure algebraic-code-excited linear prediction", June 1995.

ITU-T Recommendation G.723.1, "Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kb/s", March 1996.

"APCO/NASTD/Fed. Project 25 Vocoder Description", Telecommunications Industry Association Specifications, 1992.

D. Griffin and J.S. Lim, "Multiband Excitation Vocoder", in IEEE Trans. Acoustics, Speech, and Signal Processing, vol. ASSP-36, no. 8, pp.~1223--1235, 1988.

SPEECH MODIFICATION

T.F. Quatieri, R.B. Dunn, and D.A. Reynolds, "On the influence of rate, pitch, and spectrum on automatic speaker recognition performance", Proc. Int. Conf. on Spoken Language Processing, Beijing, China, October 2000.

T.F. Quatieri and R.J. McAulay, "Shape-Invariant Time-Scale and Pitch Modification of Speech", IEEE Trans. Acoustics, Speech, and Signal Processing, vol. 40, no. 3, pp. 497-510, March 1992.

S. Roucos and A.M. Wilgus, "High quality time-scale modification of speech", Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, pp. 493--490, Tampa, Florida, March 1985.

X. Serra, "A System for Sound Analysis/Transformation/Synthesis Based on a Deterministic Plus Stochastic Decomposition", Ph.D. Thesis, CCRMA, Department of Music, Stanford University, 1989.

J. Laroche, "Time and pitch scale modification of audio signals", In Applications of Digital Signal Processing to Audio and Acoustics, Eds. M. Kahrs and K. Brandenburg, Kluwer Academic Publishers, Boston, MA, 1998.

J. Laroche and M. Dolson, "Improved phase vocoder time-scale modification of audio", IEEE Trans. on Speech and Audio Processing, vol. 7. no. 3, pp. 323--332, May 1999.

NOISE REDUCTION IN THE SPEECH WAVEFORM

S.V. Vaseghi, Advanced signal processing and digital noise reduction, Wiley and Teubner, New York, NY, 1996.

T.F. Quatieri and R.B. Dunn, "Speech enhancement based on auditory spectral change", ICASSP02.

Removal of noise from noise-degraded speech signals, National Research Council panel. National Academy Press, 1989. Landmark study that shows the difficulty in obtaining a gain in intelligibility by noise suppression.

Speech Enhancement, Edited by J.S. Lim (Prentice Hall, 1983). Collection of papers, including standard methods of noise reduction.

S.F. Boll, ``Suppression of acoustic noise in speech using spectral subtraction", IEEE Trans. Acoust., Speech, and Signal Processing, Vol. ASSP-29, pp. 113-120, Apr. 1979.

M. Berouti, R. Schwartz, and J. Makoul, ``Enhancement of speech corrupted by additive noise", Proc. IEEE Int, Conf,Acoustic, Speech, and Signal Proc., pp. 208-211, Apr. 1979.

O. Cappe, ``Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," IEEE Trans. Speech and Audio Processing, Vol. 2, No. 1, pp. 345-349, April 1994.

Y. Ephraim and D. Malah, ``Speech enhancement using a minimum mean-square error log-spectral amplitude estimator", IEEE Trans. Acoustics, Speech, and Signal Processing, Vol. ASSP-33, No. 2, pp. 443-445, 1985.

J.H. Hansen and S. Nandkumar, ``Robust Estimation of Speech in Noisy Backgrounds Based on Aspects of the Auditory Process'', JASA, Vol. 97, No. 6, June 1995.

J.H. Hansen and M.A. Clements, ``Constrained iterative speech enhancement with application to automatic speech recognition'', IEEE Trans. on Signal Processing, Vol. 39, pp. 795-805, Apr. 1991.

M.R. Weiss, E. Aschkenasy, and T.W. Parsons, "Study and development of the INTEL technique for improving speech intelligibility," Nicolet Scientific Corp, Final Rep. NSC-FR/4023, December 1974.

J.D. Wise, J.R. Caprio, and T.W. Parks, "Maximum likelihood pitch estimation'', IEEE Trans. Acoustics, Speech, and Signal Processing, vol. ASSP-24, no. 5, pp. 418--423, October 1976.

P. Vary, ``Noise suppression by spectral magnitude estimation - mechanism and theoretical limits'', Signal Processing, vol. 8, pp. 387--400, 1985.

N. Virag, ``Single channel speech enhancement based on masking properties of the human auditory system'', IEEE Trans. on Speech and Audio Processing, vol.7, no. 2, pp. 126--137, March 1999.

NOISE REDUCTION FOR VOCODING PREPROCESSING

J. Jesper and J.H. Hansen, "Speech enhancement using a constrained iterative sinusoidal model", IEEE Trans. on Speech and and Audio Processing, vol. 9, no. 7, October 2001.

LINCOLN WORK

D. Messing, Noise suppression using spectral magnitude and phase from nonacoustic sensors, MS Thesis, MIT, August 2003.

T. F. Quatieri, D. Messing, K. Brady, W. B. Campbell, J. P. Campbell, M. Brandstein, C. J. Weinstein, J. D. Tardelli, and P. D. Gatewood, "Exploiting nonacoustic sensors for speech enhancement", Proc. Workshop on Multimodal User Authentication, Santa Barbara, CA, 11-12 December 2003.

W.M. Campbell, T.F. Quatieri, J.P. Campbell, and C.J. Weinstein, "Multimodal speaker authentication using nonacoustic sensors," Proc. Workshop on Multimodal User Authentication, Santa Barbara, CA, 2003.

K. Brady, T. F. Quatieri, W. B. Campbell, J. P. Campbell, M. Brandstein, C. J. Weinstein, "Multisensor MELPe using parameter substitution", Proc. Int. Conf. Acoustics, Speech, and Signal Processing, Montreal Canada, 2004.

NONACOUSTIC SENSORS

M. Rothenberg, "A multichannel electroglottograph," J. of Voice, vol. 6, no. 1, pp. 36-43, 1992.

M.V. Scanlon, "Acoustic sensor for health status monitoring," Proceedings of IRIS Acoustic and Seismic Sensing, vol. 2, pp. 205-222, 1998.

T. Yanagisawa and K. Furihata, "Pickup of speech signal utilization of vibration transducer under high ambient noise", Journal of Acoustical Society of Japan, Vol. 31, No. 3, pp. 213-220, 1975.